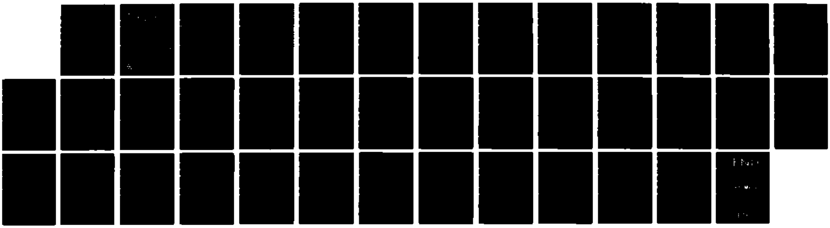
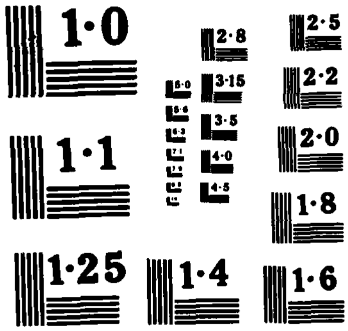
AD-A155 623 MEDIUM RANGE ATMOSPHERIC FORECAST SYSTEM: EVALUATION PLAN(U) SYSTEMS AND APPLIED SCIENCES CORP MONTEREY CA P A PETIT ET AL. FEB 85 NEPRF-CR-85-03 N00228-83-C-3101 1/1

UNCLASSIFIED F/G 4/2 NL





NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

NAVENVPREDRSCHFAC
CONTRACTOR REPORT
CR 85-03

NAVENVPREDRSCHFAC CR 85-03

AD-A155 623

MEDIUM RANGE ATMOSPHERIC FORECAST SYSTEM: EVALUATION PLAN

Pierre A. Petit, Harry D. Hamilton, Russell L. Elsberry
Systems and Applied Sciences Corporation
Monterey, CA 93940

Contract No. N00228-83-C-3101

FEBRUARY 1985

85 5 22 080

Prepared For:
NAVAL ENVIRONMENTAL PREDICTION RESEARCH FACILITY
MONTEREY, CALIFORNIA 93943-5106

QUALIFIED REQUESTORS MAY OBTAIN ADDITIONAL COPIES FROM THE DEFENSE TECHNICAL INFORMATION CENTER. ALL OTHERS SHOULD APPLY TO THE NATIONAL TECHNICAL INFORMATION SERVICE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) CR 85-03
6a. NAME OF PERFORMING ORGANIZATION Systems and Applied Sciences Corporation; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Naval Environmental Prediction Research Facility
6c. ADDRESS (City, State, and ZIP Code) 570 Casanova Avenue Monterey, CA 93940	7b. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5106
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Naval Air Systems Command; 8b. OFFICE SYMBOL (If applicable) (AIR-330)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00228-83-C-~~5106~~ 3101
8c. ADDRESS (City, State, and ZIP Code) Department of the Navy Washington, DC 20361	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 63207N; PROJECT NO. 7W0513; TASK NO.; WORK UNIT ACCESSION NO. DN293603

11. TITLE (Include Security Classification)

Medium Range Atmospheric Forecast System: Evaluation Plan

12. PERSONAL AUTHOR(S) Petit, Pierre A.; Hamilton, Harry D.; Elsberry, Russell L.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 8/83 TO 2/85	1985, February	40

16. SUPPLEMENTARY NOTATION
See also contractor report NAVENVPREDRSCHFAC CR 85-04, "Medium Range Atmospheric Forecast System: Evaluation Report" dated February 1985

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Extended range numerical prediction,
04	02		Numerical weather prediction.
			Medium range forecast

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A plan for conducting a comprehensive evaluation of the accuracy and utility -- with respect to operational Navy applications -- of a medium range (5-10 day) atmospheric forecast system is presented. Detailed evaluation procedures for verification data collection and reduction for practical skill assessment are provided.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Payne, Steven W. (contract monitor)	22b. TELEPHONE (Include Area Code) (408) 646-2326; 22c. OFFICE SYMBOL NEPRF WU 6.3-20

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

TABLE OF CONTENTS

TABLE OF CONTENTS (continued) PAGE

SECTION 1. EVALUATION OVERVIEW

1.1 Purpose of Plan. To set forth a procedure for the comprehensive evaluation of the accuracy and utility, with respect to operational Navy applications, of a medium-range atmospheric forecast system.

1.2 Background. The U.S. Navy has been preparing and applying numerical weather prediction products since the early 1960's; but, until recently, the maximum forecast period was limited to 72 hours. Computers with larger memories and smaller instruction execution times have evolved over the years and much longer forecast periods are now feasible in operational environments, such as at the Fleet Numerical Oceanography Center (FNOC) in Monterey, California. Providing the newer, more sophisticated atmospheric prediction models actually have skill in the longer range, their output would be very valuable for many Navy applications.

FNOC is the primary site for large-scale, numerical environmental prediction in the Navy. Atmospheric prediction model development for the Navy is accomplished by the Naval Environmental Prediction Research Facility (NEPRF), which is also located in Monterey. In 1981 FNOC installed a new, global atmospheric forecasting system called NOGAPS (Navy Operational Global Atmospheric Prediction System) which had been developed by NEPRF. In 1983 NEPRF upgraded NOGAPS' physics and resolution and FNOC began integrating the model to 5 days. The Navy anticipates that NOGAPS, in its present form or with some further upgrading, can produce skillful medium-range forecasts. Medium range is defined as 5 to 10 days for this plan.

The Navy has had little experience in the evaluation and use of numerical environmental predictions for periods greater than 3 days and there is little guidance on how to make operational use

of medium-range numerical forecasts. Recognizing this situation, NEPRF undertook a project to develop a procedure which could establish the relevant accuracy and operational utility of a medium-range atmospheric forecast system such as NOGAPS.

As a first task in the Medium-Range Forecast Evaluation (MRFE) project, a review was prepared to address the present accuracy and operational use of medium-range numerical forecasts - with an emphasis on Navy applications. That review (Elsberry, Hamilton and Petit, 1984) describes present levels of medium-range forecast skill (for example, at the European Centre for Medium Range Weather Forecasts (ECMWF)) and it sets forth acceptable medium-range levels of accuracy for various operationally relevant weather parameters.

Based on the aforementioned report, the second task in the MRFE project is the preparation of this plan for evaluating the likely operational worth of a medium-range forecast system. Future tasks in the MRFE project will be to conduct the evaluation in accordance with this plan, to assess the results in terms of a baseline evaluation, and to document any procedural changes which may be indicated for similar evaluations in the future.

1.3 Rationale. In the early stages of developing this Medium-Range Atmospheric Forecast Evaluation Plan, several basic requirements were identified. Briefly stated they are:

- That the evaluation be objective.
- That the results be operationally relevant and scientifically convincing.
- That the procedure be computer efficient and not be labor intensive.
- That the procedure be easily repeatable.

The reasons for these requirements and the degree to which this plan responds to them are discussed in the subsections which follow.

<u>1.3.1 Objectivity</u>. Previous Navy forecast system evaluations (for example, Wash et al, 1982) have been to a large extent subjective. Several experienced operational forecasters spent many hours examining plots of model output and evaluating the practical worth of each days forecast series relative to some alternative such as an older forecast model. Opinions were stratified by broad characterizations such as "poor" or "inferior" through "fair" or "equal" to "excellent" or "superior". Usually some objective skill measures such as root-mean square error were calculated, but seldom were these objective scores rigorously correlated with the subjective opinions.

For reasons of economy (as discussed in subsection 1.3.3) and to improve repeatability (see subsection 1.3.4) it was decided to minimize subjectivity in these medium-range forecast verification and evaluation procedures. This goal has been met. All of the field verifications will be objective and two of the three special verifications will be objective.

The one subjective part of this plan involves sensible weather forecasting and verification. For this part of the evaluation, approximately 40 hours each will be required of four forecasters who will predict operationally relevant weather elements for a set of 9 extratropical locations and one tropical ocean area. Sensible weather forecast skill scores will be correlated with other objective measures such as anomaly correlation scores. (The details of the sensible weather forecasting and verification procedure are in subsection 2.1.2.3.)

Once such a baseline measure of operational forecaster skill has been established, it is expected that strictly objective measures would suffice for future "technical evaluations" within FNOC. However, the sensible weather forecasting and verification procedures might need repeating in future "operational evaluations" in order to refine the sensible weather baseline for other areas and seasons. This would also provide a structured, organized way for forecasters in the field to gain familiarity with substantially new or improved atmospheric forecast models. As long as sensible weather parameters such as visibility and precipitation remain unanalyzed, some subjective evaluation will probably be unavoidable.

1.3.2 Operational and Scientific Relevancy. The results of this evaluation will be of interest to two distinct groups. First, the operational forecasters and their chain of command expect the evaluation to be operationally relevant. For example, they would like to know how well storm tracks are forecast and what the gale warning false-alarm rate is. Second, the forecast system developers and their sponsors expect the evaluation to include measures of skill that illustrate the capability of the dynamics or physics of the model. In particular, this group needs to know how the skill compares with earlier versions or with similar contemporary models.

To meet the needs of the first group, emphasis is placed on verifications that relate to the acceptable levels of accuracy set forth in Elsberry et al (1984) and reproduced here as Table 1-Ø1. The three special verifications (subsection 2.1.2) dealing with storm tracks, area wind warnings and sensible weather forecasting are designed to treat the parameters and measures contained in the table.

The more traditional field verifications discussed in subsection 2.1.1 will provide mean errors, standard deviations and similar statistics at four standard levels for three

WEATHER PARAMETER	ACCURACY AT FIVE DAYS	ACCURACY AT TEN DAYS
Extratropical Storm Track	200 nm avg. STE[1]	400 nm avg. STE[1]
Wind		
Sfc[2] speed	± 25%	± 50%
Sfc[2] direction	± 45 degrees	± 60 degrees
FA[3] speed	± 20%	± 40%
FA[3] direction	± 45 degrees	± 60 degrees
Temperature		
Sfc[2]	± SStdDev[4] • 0.4	± SStdDev[4] • 0.7
FA[3]	± 5°C	± 10°C
Clouds		
cover	± 25% (± 2/8)	clear or scattered/ broken or overcast
dominant type	cumuliform/mixed/ stratiform	cumuliform/mixed/ stratiform
base of dominant	low/middle/high	low/high
Precipitation	likely/possible/unlikely	likely/unlikely
amount (if likely/possible)	light/moderate/heavy	light/heavy
type (if likely/possible)	steady/mixed/showers	PNP[5]
frozen (if likely/possible)	yes/possible/no	likely/unlikely
Visibility	<3/3-6/>6 mi	PNP[5]
Waves (sea, swell & surf)	(sfc wind & geography dependent)	(sfc wind & geography dependent)

NOTES:

[1]STE is Surface Track Error; the minimum distance between forecast cyclone positions at prime synoptic times and the verifying cyclone track.

[2]Sfc is surface value at about two meters altitude.

[3]FA is free atmosphere above the planetary boundary layer.

[4]SStdDev is the Seasonal Standard Deviation.

[5]PNP means probably not predictable.

TABLE 1-01. Acceptable Levels of Accuracy.

latitudinal bands. These field statistics will permit relatively easy comparisons with earlier Navy verifications as well as with other numerical weather forecast centers and, in particular, with those of ECMWF.

<u>1.3.3 Economy and Efficiency</u>. The basic tasking for this plan states that it "must be capable of implementation utilizing currently available NAVENVPREDRSCHFAC manpower and computer resources." In subsequent discussions it was agreed that "currently available" meant that the evaluation should not significantly disrupt work on other ongoing projects or cause a substantial increase in the computer processing or rotating mass storage load on the FNOC computers.

The requirement to conserve manpower has dictated a largely hands-off, highly automated and objective evaluation approach. In addition to the time required for sensible weather forecasting and verification, as previously discussed, labor will be required to submit the various verification jobs to the computer and monitor their results, to organize the data reduction and interpretation and to prepare the evaluation report. There will also be one-time software development costs for several verification, analysis and summary programs not currently available; but, whenever possible, on-the-shelf software will be used with minimum modification. Once this software development is complete, and excluding the sensible weather portion of the evaluation, any future repetitions of this plan should not be very manpower intensive.

A most important, basic decision was to not integrate and therefore not evaluate the forecast model beyond 7 days. Beyond that, computer efficiency in terms of on-line storage, prime-time and off-time has been emphasized in that order. The number of field verifications have been limited to minimize the preparation and temporary archiving of data fields which would be of no particular operational interest. More fields would require more

space than can conveniently be made available on operational FNOC rotating mass storage. Rather than save data for eight vertical levels and 15 geographic areas, as ECMWF does and as would be convenient for diagnostic and detailed intercomparison purposes, this plan calls for verifying only four levels and five areas which is more efficient and adequate for an operationally oriented evaluation such as this. Prime-time jobs will be limited to only those absolutely necessary to ensure that required data are not lost. Detailed verification and data reduction will be scheduled at those times of the day and week when the operational load is lowest.

1.3.4 Repeatability. Because the evaluation procedure is objective, economical, efficient, and well documented, it will satisfy the requirement of repeatability. This attribute is important to those interested in documenting the changing skill of an evolving medium-range forecast system and also to those interested in making comparisons with other, perhaps new, forecast models.

Since the first, pilot execution of this plan will be for a limited period of about two months, it will be necessary to repeat this evaluation for other months and seasons. To facilitate this process and identify any needed procedural changes or enhancements prior to a repetition, a summary critique is specified in subsection 2.3.3.

1.4 Plan Summary and Schedule. This plan, which is detailed in Section 2, provides for two general types of evaluations (accuracy and utility) conducted in three phases (data collection and verification, verification data reduction, and summarization). The accuracy evaluation will be based principally on fairly traditional field verifications at four standard levels, three in the troposphere and one mostly in the lower stratosphere. Some spectral truncations and time averages that are particularly appropriate for medium-range skill

assessment will also be included. The utility evaluation will be based primarily on three special verifications of storm tracks, area wind warnings and subjective sensible weather forecasts. These evaluations will be related to the parameters and levels of accuracy set forth in Table 1-Ø1.

The beginning of the data collection and verification phase is targeted for 1 April 1984. Day 4 through Day 7 forecast fields from the ØØZ base-time run will be saved in one day increments and verifications will commence when verifying analyses become available. This phase will last 9 to 1Ø weeks, which is sufficient time to verify a complete two months of data (provided, of course, that FNOC successfully integrates NOGAPS out to 168 hours (TAU 168) every day of each week based on the ØØZ data). The collection phase will result in daily raw error statistics at grid-points, at all locations for which sensible weather forecasts are prepared, over the selected wind warning areas, and within the North Pacific storm track area.

The verification data reduction phase will commence when the collection phase completes. It will require about two calendar months, but on a computer intensive rather than labor intensive basis. In this phase the raw error statistics will be reduced to area and monthly means, standard deviations, root-mean square errors and similar measures of skill.

The summarization phase will commence near the end of the data reduction phase and will also require about two months. This time will be spent preparing summary graphics and a draft, final written evaluation report and critique. That report will provide baseline estimates of both model skill and model utility for several atmospheric levels, geographic areas and operationally relevant parameters. Lessons learned during this pilot evaluation and recommended changes for any subsequent medium-range evaluations will be included in the report.

Allowing a reasonable time for data collection and reduction and for summarizing, it is projected that the draft final report will be delivered in early October if data collection begins on 1 April 1984.

SECTION 2. EVALUATION PROCEDURES AND PRESENTATION

2.1 Data Collection and Verification. This plan requires the collection of two basic types of data: field data which are needed primarily to assess accuracy; and special verification data sets which are required primarily for utility assessment. Eighty basic data fields will be saved for verification purposes and 52 more fields will be plotted for use by the sensible weather forecasters. All of these fields are shown in Table 2-01. A further 28 fields will be derived from the basic fields and subsequently verified. These are listed in Table 2-02. In addition, synoptic reports from several locations will be collected for use in sensible weather forecasting and verification. The details of this data collection phase of the plan and the associated verification procedures are provided in the next seven subsections. The data reduction and summarization phases are then discussed at the end of this section.

2.1.1 Field Verifications. In this plan we are concerned with fields of atmospheric analysis and forecast model variables, or derived products, on the standard FNOC spherical grid. This is a 2.5 degree latitude by 2.5 degree longitude global grid and is henceforth referred to as the standard grid or simply the grid. (Note: The NOGAPS model variables are integrated (forecast) on sigma (pressure-related) surfaces using a 2.4 degree latitude by 3.0 degree longitude grid. The FNOC output software routinely interpolates and extrapolates these values to the standard grid for sea level and for all standard pressure surfaces within the model's atmosphere.)

2.1.1.1 Basic Field Verification. Table 2-01 shows with "V" notation those variables (pressure height, temperature and wind), levels (1000, 850, 500 and 200 mb) and forecast times (96, 120, 144 and 168 hours) which will be saved in standard field form. The forecast variables will be verified by computing the difference between the forecast value at a grid point and the

FNOC Spherical Field		TAUs (forecast plus-time values)								See Note(s)
ID	DESCRIPTION	0	24	48	72	96	120	144	168	
A01	Sea Level Pressure	P	P	P	PS	PS	PS	PS	PS	1
A07	Sfc Air Temp	P		P		P	P	P	P	
A15	Sfc Vapor Pres	P		P		P	P	P	P	
C00	1000 Mb Ht. (D)	V				V	V	V	V	
C10	1000 Mb Temp	V				V	V	V	V	4
C20	1000 Mb Wind E-W (u)	PVA		P		PV	PVA	PV	PVA	2,3,4
C21	1000 Mb Wind N-S (v)	PVA		P		PV	PVA	PV	PVA	
D00	850 Mb Ht. (D)	V				V	V	V	V	
D10	850 Mb Temp	VP				VP	VP	VP	VP	
D20	850 Mb Wind E-W (u)	PV				V	PV	V	PV	2,3
D21	850 Mb Wind N-S (v)	PV				V	PV	V	PV	
F00	500 Mb Ht. (D)	VP	P	P	P	VP	VP	VP	VP	4
F10	500 Mb Temp	V				V	V	V	V	
F20	500 Mb Wind E-W (u)	V				V	V	V	V	4,5
F21	500 Mb Wind N-S (v)	V				V	V	V	V	
F8B	500 Mb Vorticity	P		P		P	P	P	P	
I00	200 Mb Ht. (D)	V				V	V	V	V	
I10	200 Mb Temp	V				V	V	V	V	
I20	200 Mb Wind E-W (u)	V				V	V	V	V	5
I21	200 Mb Wind N-S (v)	V				V	V	V	V	

TABLE 2-01. Fields Required for Medium-Range Forecast Evaluation.

FNOC Spherical Field		TAUs (forecast plus-time values)								See Note(s)
ID	DESCRIPTION	0	24	48	72	96	120	144	168	
A62	Precipitation			P		P	P	P	P	
Z21	Showers only					P	P	P	P	
DEL	PBL Depth					P	P	P	P	
STB	Stratus Thickness					P	P	P	P	
STF	Stratus Frequency					P	P	P	P	

LEGEND: V - saved for field verification.
P - plotted for sensible weather forecasters use.
A - saved for special area wind warning verification.
S - saved for storm track verification.

Note 1: 12 hour intermediate TAUs also required for SEIS.

Note 2: Single plot in ddff format.

Note 3: Verify dd and ff separately.

Note 4: Several derived fields will also be plotted and/or verified (see text).

Note 5: Verify dd, ff and vector winds separately.

TABLE 2-01. Fields Required for Medium-Range Forecast Evaluation (continued).

corresponding analyzed (TAU 0) value four to seven days later. All grid points on the standard grid from 80 North to 80 South will be verified for all fields, except only winds will be verified between 20N and 20S.

Anomaly correlation scores for the pressure height and temperature fields will be calculated at all four levels. The anomalies will be derived from the FNOC seven year climatology (June 1974 through May 1981).

In the tropics only (20N to 20S) and at 1000 and 200 mb only, threshold change statistics and the persistence errors for the wind fields will also be computed. This will be done by comparing the base-time analysis of wind (direction, magnitude and vector values) with the verifying analyses 4 to 7 days later.

2.1.1.2 Derived Field Verification. In addition to verifying the basic fields, 28 special fields will be derived from standard 1000 and 500 mb fields. These derived fields are listed in Table 2-02. The spectral truncations of the forecast 500 mb heights will be verified with the appropriate analysis (TAU 0) field. The time averages (same base time, different verifying times) and lagged averages (different base times, same verifying time) of height will be verified against the appropriate untruncated or truncated analysis. All of the day five (TAU 120) time and lagged ranges will be verified by comparison with analyzed grid point values to determine which and how many of the analyzed points are within the "forecast" (unaveraged) range.

The derived fields, like the basic fields, will be verified on the standard grid from 80 North to 80 South, except only winds will be verified between 20N and 20S. Spectral truncations will be obtained from fast Fourier transforms performed in the east-west direction only.

Mb Level	Derived Field Description	TAU				
		0	96	120	144	168
1000	Time Temperature Range			←V→		
1000	Time Wind Direction Range			←V→		
1000	Time Wind Magnitude Range			←V→		
1000	Lagged Temperature Range			V*		
1000	Lagged Wind Direction Range			V*		
1000	Lagged Wind Magnitude Range			V*		
500	Spectral Height Truncation Waves 1 thru 3	A	V	V	V	V
500	Spectral Height Truncation Waves 4 thru 9	A	V	V	V	V
500	Time Average Height All Waves			←V→		
500	Time Average Height Trunc. Waves 1 thru 3			←V→		
500	Time Average Height Trunc. Waves 4 thru 9			←V→		
500	Lagged Average Height All Waves			V*		
500	Lagged Average Height Trunc. Waves 1 thru 3			V*		
500	Lagged Average Height Trunc. Waves 4 thru 9			V*		
500	Time Temperature Range			←V→		
500	Time Wind Direction Range			←V→		
500	Time Wind Magnitude Range			←V→		
500	Lagged Temperature Range			V*		
500	Lagged Wind Direction Range			V*		
500	Lagged Wind Magnitude Range			V*		

LEGEND:
- A - derived for use as a verifying analysis.
- V - derived from verification.
- V* - derived from the same calendar day and time data from the three most recent daily forecast runs.
- ←V→ - derived from the three TAUs indicated, all from same run.

TABLE 2-02. Non-standard Fields to be Derived.

This set of derived field verifications will test the medium-range model's skill in forecasting the longer waves and its ability to predict the probable ranges (or bounds) of important parameters. Such skill would not be readily apparent from the more traditional, basic field verifications.

2.1.2 Special Verifications.

2.1.2.1 Storm Tracks. The degree to which a forecast system can successfully predict the tracks and central values of low pressure systems is a very good gauge of its practical operational value. This was recognized by NEPRF when they developed their Systematic Error Identification System (SEIS), the heart of which is a Vortex Tracking Program (VTP). The VTP is based on the work of Williamson (1979a, b and 1981) and the SEIS is summarized by Harr et al (1983). SEIS tracks up to five pressure centers (highs only or lows only, but not both in a single run) and then correlates analyzed centers with forecast centers. Various errors relating to the forecast track and center amplitude are calculated and the resulting raw verification data may be displayed and statistically summarized in various ways. For this evaluation the planned operational, short-range application of SEIS at FNOC will be modified and extended to provide for the calculation of medium-range verification data. In addition, two special storm track verifications will be conducted. These will measure the models skill in forecasting the five day mean storm track rather than its skill in forecasting the tracks of individual centers as SEIS does.

2.1.2.1.1 Systematic Error Identification System (SEIS). SEIS is already running within a North Pacific area on low pressure centers only. The SEIS area was chosen by NEPRF and FNOC to capture data on the very Navy-relevant North Pacific storm track.

Details concerning the running of SEIS in support of this evaluation plan are:

- To be run over the NEPRF/FNOC defined North Pacific area on low pressure centers only for at least one month and not more than two months.

- Since SEIS requires 12 hour track continuity to be reliable and is computationally a long job, it will be run on the following three sets:

  - Set A: TAUs 00, 12, 24, 36, 48 and 60
  - Set B: TAUs 72, 84, 96, 108 and 120
  - Set C: TAUs 132*, 144, 156* and 168

  *Note: TAU 132 and 156 fields are not written by NOGAPS, but will be created by averaging adjacent TAUs.

- Set A will be run routinely by FNOC/NEPRF SEIS personnel for both the 00Z-based and the 12Z-based forecast fields.

- Sets B and C will be separately run by medium-range evaluation personnel for the 00Z-based forecast fields only.

- Set B and C job streams will be based on the NEPRF Set A job stream which takes Northern Hemispheric polar stereographic 63 x 63 gridded sea level pressure fields through Field Separation (zonal mean removed and resultant "D" fields written to intermediate file) and Vortex Tracking (output is an extended "Raw Verification Data" file). Job stream modifications will include spherical-to-polar conversion and field averaging for TAUs 132 and 156.

- Sets B and C will be run in date-time-TAU order; each set a separate job to be completed through Vortex Tracking within 8 days of base date-time.

- The Error Statistics program which uses the Raw Verification Data as input and writes the Raw Error Statistics file will be run when convenient by NEPRF SEIS personnel and at least once within 30 days following the data collection period.

The reduction of SEIS medium-range error statistics is discussed in subsection 2.2.2.1.

2.1.2.1.2 Mean Storm Track Verification. Five day mean storm tracks (forecast and analyzed) will be constructed by two different methods for this evaluation. In the first method, five day lowest-grid-point-value composites centered on forecast Day 5 (TAU 120) will be constructed from the SEIS vortex tracking program's elliptical function lows only output. In the second method, the same composites will be separately constructed using standard sea level pressure analysis and forecast fields from which the zonal mean (ZM) will be removed and in which all greater than ZM grid point values will be set to one. The first method is economical when and over those areas where SEIS is run, but it would be computationally expensive to do otherwise. The second method is computationally reasonable for any time and area. For this evaluation the two methods will be compared and if, as expected, they are roughly equivalent verification tools, only one would be chosen for future evaluations.

In both methods, each five day mean forecast field will subsequently be compared with the verifying five day composite analysis field and the five day mean storm track correlation coefficient will be calculated.

2.1.2.2 Area Wind Warnings. Hazardous winds (and the hazardous seas associated with such winds) are of continuing concern to the Navy's operating forces. This special verification will permit assessing model skill in predicting gale force winds over a limited ocean area 5 to 7 days in advance. It is the magnitude of the forecast boundary layer winds which will be verified.

Details concerning this portion of the evaluation follow.

- Size of areas: 7.5 degrees latitude by 15 degrees longitude at even 2.5 degree intersections of the standard grid.

- Location of areas:

  52.5-60.0N/15.0-30.0W and 175E-170W
  42.5-50.0N/35.0-50.0W and 155-170E
  32.5-40.0N/60.0-75.0W and 140-155E
  22.5-30.0N/15.0-30.0W and 150-165W

  (Eight areas equally divided between Atlantic and Pacific, all elongated in the east-west direction and favoring the preferred storm tracks.)

- Forecast Variables:

  Gale force or stronger winds in area on day 5 - yes or no
  Gale force or stronger winds in area on day 7 - yes or no

- Selection Criteria:
  - Yes if ten or more of the 28 grid points within or on the perimeter of the area are forecast to have winds in excess of 32 knots at 00Z.
  - No if less than ten grid points meet the above criterion.

  (Ten grid points define about 25 percent of the area if contiguous. This less than 50 percent criterion is

conservative and provides some allowance for any phase speed error.)

- <u>Scoring Criteria (verification):</u>
  - If gale forecast was yes:

    $\geq$ 10 points $\geq$ 30 knots = 100%
    9 points $\geq$ 30 knots = 90%
    8 points $\geq$ 30 knots = 80%
    etc.
    0 points $\geq$ 30 knots = 0%

  - If gale forecast was no:

    $\leq$ 9 points $\geq$ 36 knots = 100%
    10 points $\geq$ 36 knots = 90%
    11 points $\geq$ 36 knots = 80%
    etc.
    $\geq$ 19 points $\geq$ 36 knots = 0%

  (The 30 and 36 knot yes/no verification criteria are purposely in the "model's" favor.)

Verification scores will be computed regularly throughout the data collection phase. The area sizes and/or locations could be changed for future evaluations and the selection or scoring criteria could be tuned at the same time.

<u>2.1.2.3 Subjective Sensible Weather Forecasts</u>. The third special verification will evaluate the usefulness of the forecast model's output as guidance material for forecasters preparing medium-range sensible weather predictions. In addition to the standard pressure-height, temperature and wind fields, several less familiar boundary layer forecast fields will be made available for forecaster consideration.

Four persons (civil service and contractor) with operational Navy weather forecasting experience will participate. (For

future evaluations, active duty military forecasters at distant Navy installations could be used.) Three of these forecasters will each be assigned a geographic area from among NEPAC (20-70N, 180-120W), NWLANT (20-70N, 80-35W) and NELANT (20-70N, 35W-10E). They will be asked to select three well exposed island or coastal synoptic reporting stations within their area. These stations are to be extratropical, well separated and known to have reliable weather reports. Some forecaster familiarity with the selected area and stations is desired. This experience level as well as any previous exposure to the forecast model output will be documented. The fourth forecaster will be assigned responsibility for forecasting the position of the intertropical convergence zone (ITCZ) over the equatorial Western Pacific (from 100E to 180E). Each forecaster will prepare a Day 5 and a Day 7 forecast for each of his three extratropical reporting stations or for his tropical area on each Tuesday and Thursday for nine weeks. The first week will be a start-up period during which verification scores will not be recorded. A supernumerary forecaster will be available to substitute in cases of absence and any substitutions will be documented.

Figure 2-01 shows the forecast elements and the form which will be used for extratropical forecasting. Forecasters will be limited to two hours each forecasting day in which to study the latest guidance material and fill out their three forecast forms. All will have access to the same guidance and will be asked to prepare their forecasts without reference to any other material. Guidance material made available will consist of those NOGAPS fields identified by "P" notation in Table 2-01 plotted on Northern Hemisphere Polar Stereographic backgrounds and the most recent synoptic reports through 06Z on the forecasting day for each of the forecast stations. Field plots and synoptic reports will be centrally displayed each Monday through Thursday for the nine-week period. Verification will be done by comparing the forecast values for each parameter with the station's observations 5 or 7 days later and assigning a numerical score.

Tue
Thu

Forecaster's Name ________ Block/Station Nr. ________ Day / Date / Mo. ________

Forecast Parameter	Day Five			Day Seven		
	Fcst	Obs	Score	Fcst	Obs	Score
Winds $\geq$ 25 kt (18-06Z) - yes/no						
Avg Sfc Wind Spd (18-06Z) - kt						
Avg Sfc Wind Dir (18-06Z) - quad't[1]						
Sfc Air Temp (00Z) - deg. F						
Avg Cloudiness (18-06Z) - 8ths						
Lowest Cloud Base (18-06Z) - 100s						
Precip Expected (18-06Z) - yes/no						
and only if yes: Rain or Shower - R/S						
Frozen - yes/no						
Sfc Visibility $\leq$ 3 Mi (00Z) - yes/no						

[1]Quadrants are: 1-N, 2-NE, 3-E, 4-SE, 5-S, 6-SW, 7-W, 8-NW

FIGURE 2-01. Medium-range Evaluation Sensible Weather Forecast Form

The criteria for this process are shown in Table 2-03. Scrutiny of that table will show that flexibility in the forecaster's favor has been provided. For example, one report of precipitation will verify a "yes" forecast, but two are required to fail a "no" forecast. Similarly, when forecasting visibility less than or equal to three miles, the forecaster gets credit for any report less than four miles if he forecasts "yes" and does not get penalized until two and one quarter miles or less is observed when the forecast is "no". As was true with area wind warnings, forecast parameters and scoring criteria could be easily changed or tuned for future evaluations.

The tropical ITCZ forecaster will use plots of the TAU 00, 120 and 168 wind fields at 1000 and 850 mb as well as the latest, appropriate satellite depiction to prepare a Day 5 and Day 7 forecast of ITCZ latitude in whole degrees at each five degrees of longitude from 100E through 180E. Verification will be done by visually determining the verifying location of the (or the dominant) ITCZ to the nearest whole degree of latitude and then recording the difference in degrees between this and the forecast latitude for each five degrees of longitude. (Note: if the required satellite imagery is not readily available, or if the early ITCZ forecasting results are particularly discouraging; a request will be made to the NEPRF evaluation coordinator for permission to reassign the ITCZ forecaster as a fourth extratropical forecaster with responsibility for NWPAC (20-70N, 120-180E).)

2.2 Verification Data Reduction. This second phase of the plan is to reduce the large amount of raw verification data in ways that will permit meaningful comparisons of skill between levels, TAUs, areas and parameters, and, at least to some extent, comparison with other model verifications. The actual comparisons and their interpretation comprise the third phase of the plan which is discussed at the end of the section.

PARAMETER	SCORING RULES
Winds $\geq$ 25 Ks	If fcst yes: any 18-06Z report $\geq$ 20 scores 1, otherwise 0 If fcst no: any 18-06Z report $\geq$ 30 scores 0, otherwise 1
Avg Sfc Wind Spd	Fcst value - avg of 18-06Z reports = score (with sign)
Avg Sfc Wind Dir	If fcst quadrant = geometric avg of 18-06Z score 1, otherwise 0
Sfc Air Temp	Fcst value - avg of 21Z-03Z reports (deg F) = score (with sign)
Avg Cloudiness	Fcst value - avg of 18-06Z total reported clouds (N) = score (with sign)
Lowest Cloud Base	Fcst value - avg of 18-06Z reported bases (h as "plotted") = score (with sign)
Precip Expected	If fcst yes: any 18-06Z ww>49 or any 21-06Z W>4 scores 1, otherwise 0 If fcst no: more than one 18-06Z ww>49 or 21-06Z W>4 scores 0, otherwise 1
Rain or Shower	If precip score = 0 this score is N (null) If precip score = 1 and precip fcst was no, this score is N Otherwise: 18-06Z avg of ww>49 and (Wx10)>40 = PTA Then: If fcst was rain and PTA<77 score 1 If fcst was shower and PTA>78 score 1 Otherwise score 0
Frozen	If precip score = 0 this score is N (null) If precip score = 1 and precip fcst was no, this score is N Otherwise if fcst yes: any 18-06Z ww 56-57, 66-79, 83-90 or 93-97 or 99 or any 21-06Z W of 7 scores 1, otherwise 0 if fcst no: less than two of the above scores 1, otherwise 0
Sfc Visibility	If fcst was yes and any 21-03Z coded VV<56 ($\underset{\sim}{<}$4 mi) score 1 If fcst was no and all 21-03Z coded VV>35 ($\underset{\sim}{>}$2¼ mi) score 1 Otherwise score 0

TABLE 2-03. Sensible Weather Forecast Scoring Criteria

2.2.1 Field Data Reduction. For this evaluation the field verification data will be analyzed for each month and both months combined, for each of days 4 through 7, for a set of geographic areas in terms of several objective scores. Table 2-04 lists the five major areas, the seven subareas and related information.

AREA	COORDINATES	NR. OF STD. GRID POINTS
Northern Hemisphere	20N-80N,60E-0-57.5E	3600
Tropics	20N-20S,60E-0-57.5E	2448
Tropical N. Atlantic	00N-20N,20-90W	261
Trop. W. Pacific	20N-20S,100E-180	1241
N. Indian Ocean	00N-20N,60-100E	153
Southern Hemisphere	20S-80S,60E-0-57.5E	3600
North Pacific	70N-20N,120E-120W	1029
Northwest Pacific	70N-20N,120E-180	525
Northeast Pacific	70N-20N,180-120W	525
North Atlantic	70N-20N,80W-10E	777
Northwest Atlantic	70N-20N,80W-35W	399
Northeast Atlantic	70N-20N,35W-10E	399

TABLE 2-04. Field Data Areas

The North Pacific area approximates the corresponding SEIS areas. The two equal-sized subsets of each ocean basin will be recognized as the broad sensible weather forecasting areas. There is a four to three size ratio between corresponding Pacific and Atlantic areas. For area means (see below), a cosine weighting function will be used to compensate for the decreasing grid distance along latitudes as one proceeds from the equator to the pole.

Except as noted, the following objective scores will be computed for the basic fields for all areas, levels and variables:

- mean error of forecast (only the wind in the tropics)
- mean error of persistence (wind only, tropics only)
- root-mean-square error of forecast (only the wind in the tropics)
- root-mean-square error of persistence (wind only, tropics only)
- standard deviation of forecast error (only the wind in the tropics)
- standard deviation of persistence error (wind only, tropics only)
- standard deviation of verifying anomaly (heights and temperatures only, not in tropics)
- anomaly correlation of forecast (heights and temperatures only, not in tropics)

These scores will be computed using the following expressions:

$$1/n\ \Sigma\ (F-A_V) = (\overline{F-A_V}) = \text{mean error of forecast}$$

$$1/n\ \Sigma\ (A_O-A_V) = (\overline{A_O-A_V}) = \text{mean error of persistence}$$

$$\sqrt{1/n\ \Sigma\ (F-A_V)^2} = \text{rmse of forecast}$$

$$\sqrt{1/n\ \Sigma\ (A_O-A_V)^2} = \text{rmse of persistence}$$

$$\sqrt{1/n\ \Sigma\ [(F-A_V)-(\overline{F-A_V})]^2} = \text{standard deviation of forecast error}$$

$$\sqrt{1/n\ \Sigma\ [(A_O-A_V)-(\overline{A_O-A_V})]^2} = \text{standard deviation of persistence error}$$

$$\sqrt{1/n\ \Sigma\ [(A_V-C)-(\overline{A_V-C})]^2} = \text{standard deviation of verifying anomaly}$$

$$\frac{\Sigma\ \{[(F-C)-(\overline{F-C})]\ [(A_V-C)-(\overline{A_V-C})]\}}{\sqrt{\Sigma\ [(F-C)-(\overline{F-C})]^2 \Sigma [(A_V-C)-(\overline{A_V-C})]^2}} = \text{anomaly correlation for forecast}$$

where:

A_O = initial analysis
A_V = verifying analysis
F = forecast
C = monthly climatology
n = number of gridpoints in the verification area
F-C = predicted anomaly
A_V-C = verifying anomaly
F-A_V = forecast error
A_O-A_V = persistence error

(overbar) $\overline{\qquad}$ = area mean

Vector wind errors are calculated in wind component form as follows:

$$\vec{V}_{mean} = \sqrt{[u_{mean}]^2 + [v_{mean}]^2}$$

$$\text{rmse}(\vec{V}) = \sqrt{[\text{rmse}(u)]^2 + [\text{rmse}(v)]^2}$$

$$\text{stdv}(\vec{V}) = \sqrt{[\text{stdv}(u)]^2 + [\text{stdv}(v)]^2}$$

Vector wind errors will be calculated at 200 mb in all areas and at 1000 mb in the tropics. Scalar wind direction and magnitude errors will be separately calculated for all areas and levels.

In the tropics as a whole and in its three subareas separately, ability to forecast substantial change in vector wind at 1000 and 200 mb will be assessed at each grid point. Substantial change thresholds will be set at 10 and 25 kt for the two levels respectively. Contingency tables will then be used to compute error reduction and forecast bias statistics.

The derived fields listed in Table 2-02 will be verified for the two hemispheres and two ocean basins only. These verifications will be in terms of mean error of forecast, rmse of forecast and standard deviation of forecast error for the spectral truncations and averages. The derived range fields will be verified in terms of area-weighted percent of points within the specified range at 00Z on Day 5.

<u>2.2.2 Special Data Reduction</u>. The special verification data will be consolidated for each month and for both months combined as specified in the following three subsections.

<u>2.2.2.1 Storm Track Data Reduction</u>. The Raw Error Statistics file from SEIS will be used to derive the following measures of skill for the Northern Pacific SEIS area for Days 4-7 inclusively:

- means and standard deviations of the forecast error (the distance in nautical miles between the forecast and verifying positions).

- means and standard deviations of the track error (the shortest distance in nautical miles between the forecast position and the verifying storm track).

- means and standard deviations of the timing error (the hourly difference between the verifying position and the position on the verifying track lying closest to the

forecast position).

- means and standard deviations of the central pressure (amplitude) error (the difference in mb between the forecast and verifying central pressures).

The daily correlation coefficients derived from both methods of mean storm track verification will be assembled and their separate means and standard deviations will be calculated for each month and for both months combined.

<u>2.2.2.2 Area Wind Warning Data Reduction</u>. The means and standard deviations of the daily area wind warning scores will be computed for Day 5 and Day 7 only, for each individual warning area, for the two ocean basins and for the Northern Hemisphere. The same information will also be computed for "yes" only and "no" only cases.

<u>2.2.2.3 Sensible Weather Forecast Data Reduction</u>. The means and standard deviations for each forecast parameter will be computed for Day 5 and Day 7, for each forecast point, for all three points in an area combined, for each ocean basin and for the hemisphere. "Yes" and "no" gale, precipitation and visibility forecasts will be considered separately. Means and standard deviations of the ITCZ forecast error in degrees will be computed for each five degrees of longitude and for all seventeen longitudes combined for Day 5 and Day 7.

<u>2.3 Summarization</u>. This the final evaluation phase is concerned with assessing and reporting the fundamental accuracy and basic utility of the forecast system. As a part of this process, lessons learned in the course of the evaluation will be documented.

<u>2.3.1 Accuracy Assessment</u>. That portion of the report which deals with accuracy will be based primarily on summary statistics

resulting from field verifications. The accuracy as a function of forecast interval and significant variations in skill from level-to-level, month-to-month and area-to-area will be described. Special attention will be given to:

- identifying differences between wavelength spectrums.

- relating anomaly correlation scores to mean and root-mean-square errors and standard deviations.

- assessing the relative merits of time and lagged averages.

- assessing the relative merits of time and lagged ranges.

- assessing the ability of the model to forecast significant vector wind changes in the tropics.

The precise tabular and geographic displays to be included in the report will depend on the results. The following is a sample of the types of displays which will be considered:

- Scatter diagram for forecast time x and level y; for example, the Day 5 (TAU 120) anomaly correlation of 500 mb height vs the standard deviation of height error.

- Weekly averages of skill measure m for variable z at level y; for example, temperature correlation at 850 mb vs week.

- Skill measure m vs forecast time x; for example, mean 500 mb height error in wave numbers 1 through 3 vs forecast days 4 through 7.

<u>2.3.2 Utility Assessment</u>. The usefulness of the medium-range forecasts will be more complicated to assess than accuracy. There are no universal measures or standards of utility. Usefulness is not only situation dependent (a model's ability to

skillfully forecast frost is not very valuable at Guam!) but also subjective if one is to consider the guidance value of a model's output. And the latter can depend nearly as much on form as on substance. A wind field can be one person's vorticity and another person's jet stream. Five wind fields can be one person's mean and another person's range.

Because of such complications, this portion of the evaluation will be done in three parts:

- First, based on the special verification data analysis, the extent of the forecast system's ability to meet the accuracy criteria of Table 1-01 will be assessed.

- Second, the results of the special verifications will be compared in key instances to the measures of skill used for accuracy assessment. For example, anomaly correlation and "range of value" scores will be compared with sensible weather forecasting skill.

- Third, forecaster opinion as to value of the guidance material <u>and</u> as to the practical meaning of the special verification results will be documented by questionnaire and summarized.

The precise form for this portion of the final report will also depend on the results. Skill score tables which summarize results by month, TAU and area will certainly be included. Scatter diagrams which relate special verifications to more standard measures of skill may also be expected; for example, forecaster skill vs anomaly correlation.

<u>2.3.3 Critique</u>. This portion of the evaluation report will relate the evaluation procedures used to the results obtained and make recommendations as to changes that might benefit future evaluations. Computer time allocations, field data availability,

special verification criteria and forecaster time budgets are examples of what will be considered in a "lessons learned" sense. It is expected that suggestions made in this portion of the report will result in effective evaluation procedures that are less labor intensive than this pilot plan and which can be followed for other months, seasons and models.

REFERENCES

Elsberry, R.L., H.D. Hamilton and P.A. Petit, 1984: Accuracy and Operational Use of Medium an Extended-Range Numerical Atmospheric Forecasts: A Review. NAVENVPREDRSCHFAC Contract Report (number not yet assigned). Systems and Applied Sciences Corporation, 570 Casanova Avenue, Monterey, California, 82 pp.

Harr, P.A., T.L. Tsui and L.R. Brody, 1983: Model Verification Statistics Tailored for the Field Forecaster. Eighth Conference on Probability and Statistics in Atmospheric Sciences, American Meteorological Society, Preprint Volume, 177-180.

Wash, C.H., L. Brody, J. Burkes, K. Freeman, S. Sandgathe, S. Silberberg, R. Toll and F. Williams, 1982: The NOGAPS product Evaluation Panel Report. Naval Postgraduate School Tech. Report prepared for Fleet Numerical Oceanography Center, Monterey, California, 55 pp.

Williamson, D.L., 1979a: Scale separation by vortex representation. Proc. 13th Stanstead Seminar, Dept. of Meteoro., McGill Univ., Montreal, Canada, 111-123.

Williamson, D.L., 1979b: Vortex representation for verification. Preprint AMS 4th Conf. on Num. Wea. Pred., Silver Spring, MD 151-154.

Williamson, D.L., 1981: Storm track representation and verification. Tellus 33, 513-530.

DISTRIBUTION

COMMANDING OFFICER
NORDA
NSTL, MS 39529

COMMANDING OFFICER
FLENUMOCEANCEN
MONTEREY, CA 93943-5105

COMMANDING OFFICER
NAVWESTOCEANCEN
BOX 113
PEARL HARBOR, HI 96860

COMMANDING OFFICER
NAVEASTOCEANCEN
MCADIE BLDG. (U-117)
NAVAL AIR STATION
NORFOLK, VA 23511

COMMANDING OFFICER
NAVPOLAROCEANCEN
NAVY DEPT.
4301 SUITLAND RD.
WASHINGTON, DC 20390

NAVAL POSTGRADUATE SCHOOL
METEOROLOGY DEPT.
MONTEREY, CA 93943

COMMANDER (2)
NAVAIRSYSCOM
ATTN: LIBRARY (AIR-723D)
WASHINGTON, DC 20361-0001

COMMANDER
NAVAIRSYSCOM (AIR-330)
WASHINGTON, DC 20361-0001

USAFETAC/TS
SCOTT AFB, IL 62225

DIRECTOR (12)
DEFENSE TECH. INFORMATION
CENTER, CAMERON STATION
ALEXANDRIA, VA 22314

LABORATORY FOR ATMOS. SCI.
NASA GODDARD SPACE FLIGHT CEN.
GREENBELT, MD 20771

EUROPEAN CENTRE FOR MEDIUM
RANGE WEATHER FORECASTS
SHINFIELD PARK, READING
BERKSHIRE RG29AX, ENGLAND

END

FILMED

7-85

DTIC